DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 38, ARTICLE 15, PAGES 335–372
PUBLISHED 25 JANUARY 2018
http://www.demographic-research.org/Volumes/Vol38/15/
DOI: 10.4054/DemRes.2018.38.15

Research Article

Global estimation of neonatal mortality using a Bayesian hierarchical splines regression model

Monica Alexander

Leontine Alkema

Contents

Demographic Research: Volume 38, Article 15
Research Article

Global estimation of neonatal mortality using a Bayesian hierarchical splines regression model

Monica Alexander[1]

Leontine Alkema[2]

Abstract

BACKGROUND
In recent years, much of the focus in monitoring child mortality has been on assessing changes in the under-5 mortality rate (U5MR). However, as the U5MR decreases, the share of neonatal deaths (within the first month) tends to increase, warranting increased efforts in monitoring the neonatal mortality rate (NMR) in addition to the U5MR.

OBJECTIVE
Data on neonatal deaths comes from a range of sources across different countries, with the amount of data available and the quality of data varying widely. Our objective in estimating the NMR globally is to combine all data sources available to obtain accurate estimates, be able to project mortality levels, and have some indication of the uncertainty in the estimates and projections.

METHODS
We present a new model for estimating the NMR for countries worldwide, using a Bayesian hierarchical model framework.

CONTRIBUTION
Our modeling approach offers an intuitive way to share information across different countries and time points, and incorporates different sources of error into the estimates. It also improves on previous modeling approaches by allowing for trends observed in NMR to be more driven by the data available, rather than trends in covariates.

[1] University of California, Berkeley, USA. E-Mail: monicaalexander@berkeley.edu.
[2] University of Massachusetts, Amherst, USA.

1. Introduction

When evaluating a country's progress in reducing child mortality, it is important to obtain accurate estimates, be able to project mortality levels, and have some indication of the uncertainty in the estimates and projections. In practice, obtaining reliable mortality estimates is often most difficult in developing countries where mortality is relatively high, well-functioning vital registration systems are lacking, and the data that is available is often subject to large sampling errors and/or of poor quality. This situation calls for the use of statistical models to help estimate underlying mortality trends.

In recent years, much of the focus in monitoring child mortality has been on assessing changes in the under-5 mortality rate (U5MR), which refers to the number of deaths before the age of 5 per 1,000 live births. The focus was driven by Millennium Development Goal (MDG) 4, which called for a two-thirds reduction in under-5 mortality between 1990 and 2015. A report on MDG progress released in 2015 by the United Nations showed that, although this target was not met in most regions of the world, notable progress has been made (UN 2015). The global U5MR is less than half of its level in 1990, and despite population growth in developing regions, the number of deaths of children under 5 has declined. Reducing the U5MR continues to be a priority as part of the Sustainable Development Goals (SDG), which replaced the MDGs in 2015. Goal 3 of the SDG includes reducing the U5MR to at least as low as 25 deaths per 1,000 live births in all countries by 2030 (UN 2017).

As the U5MR decreases, the share of neonatal deaths, i.e., deaths occurring in the first month, tends to increase. Globally, the estimated share of under-5 deaths that were neonatal in 2015 was 45%, a 13% increase from 1990 (IGME 2015). Indeed, in most regions of the world, the majority of under-5 deaths are neonatal; for example, the share is 56% in developed regions, 51% in Latin America and the Caribbean, and 54% in Western Asia. The share is still less than 50%, however, where the U5MR is relatively high; for instance, in sub-Saharan Africa the share is only 34%.

The neonatal equivalent to the U5MR is the neonatal mortality rate (NMR), which is defined as the number of neonatal deaths per 1,000 live births. The increasing importance of neonatal deaths in child mortality has warranted increased efforts in monitoring NMR in addition to the U5MR (e.g., Bhutta et al. 2010; Lawn, Cousens, and Zupan 2004; Lozano et al. 2011). Goal 3 of the SDG explicitly includes a neonatal target, with the aim to reduce the NMR to at least as low as 12 deaths per 1,000 live births in all countries by 2030 (UN 2017).

The United Nations Inter-Agency Group for Child Mortality Estimation (IGME) publishes estimates of NMR for all 195 UN member countries (IGME 2015), and these estimates are used to monitor global levels and trends in NMR over time. Until 2014, IGME used a statistical model to obtain estimates for countries without high-quality vital registration data that uses the U5MR as a predictor (Oestergaard et al. 2011). While the

method has worked well to capture the main trends in the NMR, it has some disadvantages. Most notably, trends in NMR within a country are driven by the U5MR trends, rather than being specifically influenced by the NMR data.

In this paper, we present a new model for estimating the NMR for countries worldwide, which overcomes some of the concerns with the previous IGME NMR model. We use a penalized splines regression model within a Bayesian hierarchical framework to estimate and project the NMR and to obtain uncertainty around these estimates and projections. In the model, the relationship between NMR and U5MR is used to inform estimates, and the spline regression model is used to capture country-specific trends. From the point of view of modeling mortality levels across countries, a Bayesian approach offers an intuitive way to share information across different countries and time periods, and a data model can incorporate different sources of error into the estimates.

Increases in computational speed as well as the development of suitable numerical methods has enabled a more widespread use of a Bayesian approach in many fields, including population estimation and forecasting (e.g., Alkema and New 2014; Bijak and Bryant 2016; Girosi and King 2008; Raftery et al. 2012; Schmertmann et al. 2014). The method presented in this paper has similarities to approaches used to estimate other global health indicators, including the U5MR (Alkema and New 2014; You et al. 2015), maternal mortality (Alkema et al. 2016), cause-specific mortality (Foreman et al. 2012), and contraceptive prevalence (Alkema et al. 2013). In this application, the proposed Bayesian model is flexible enough to be used to estimate the NMR in any country, regardless of the amount and sources of data available. Results were produced for 195 countries for at least the years 1990–2015, which covers the MDG period of interest, using a dataset with almost 5,000 observations from various data sources.

The remainder of the paper is structured as follows: We summarize the dataset and model in the next two sections. Some key results are then highlighted, including model validation results, followed by a discussion of the work and possible future avenues. The Appendix provides additional details about the model.

2. Data

There is large variability in the availability of data on neonatal mortality. Broadly there are three main data sources for the NMR: vital registration (VR) systems; sample vital registration (SVR) systems; and survey data. Data for a particular country may come from one or several of these sources, and the source type may vary over time. Table 1 summarizes the availability of data by source type.

2.1 Source types

Data from VR systems is derived directly from the registered births and deaths in a country. The observed NMR for a particular country and year is the number of registered deaths within the first month divided by the number of live births. Because VR data is based on the records from the whole population, it is usually high quality compared to other sources. Most developed countries have VR data available. SVR systems refer to vital registration statistics that are collected on a representative sample of the broader population.

NMR observations can also be derived from data collected in surveys, if women are asked to list a full history of all births (and possible deaths) of their children. A retrospective series of NMR observations can then be derived using the birth histories. A total of 72% of the survey data series contained in the database (Table 1) have microdata available, and it is possible to estimate the sampling error associated with each of the observations. For the remaining 28%, data comes from summary reports and preliminary releases; as such there is not enough information to calculate the sampling errors from the data. For these series, values for sampling errors are imputed (see Section 3.4.2). All mortality rates, ratios of mortality rates, and corresponding standard errors were calculated from the survey microdata using the software CMRJack (Pedersen and Liu 2012), a software package that produces mortality estimates and standard errors for surveys with complete birth histories or summary birth histories. Estimates are obtained based on the methodology outlined in Pedersen and Liu (2012). The retrospective time period covered by mortality estimates is optimized to capture short-term fluctuations while still ensuring that the estimates have a coefficient of variation of less than 10%.

The majority of survey data comes from Demographic and Health Surveys (DHS) (Table 1). The category 'other DHS' refers to non standard DHS, that is, Special Interim and National DHS, Malaria Indicator Surveys, AIDS Indicator Surveys, and World Fertility Surveys (WFS). National DHS are surveys in DHS format that are run by a national agency, rather than the external DHS agency. The Multiple Indicator Cluster Survey (MICS), developed by UNICEF in 1990, was originally designed to address trends in goals from the World Summit for Children, and has since focused on assessing progress toward the relevant MDG indicators. The 'Other' category includes surveys such as the Pan-Arab Project for Family Health and the Reproductive Health Surveys.

2.2 Data availability

Data availability varies by country and by year. For most developed countries, a complete time series of VR data exists. For other countries with VR data, the time series is often incomplete and is supported by other sources of data. Of the 105 countries where VR data is available, 44 countries have incomplete VR time series. For some smaller countries

with VR data, observations were combined to avoid issues with erratic trends due to large stochastic variance (see the Appendix for more details). SVR data is available for only Bangladesh, China, and South Africa. Most developing countries have no vital registration systems and so observations of the NMR are derived entirely from surveys. A total of 12 countries had no available data.

In terms of data inclusion, we follow the same inclusion/exclusion rules as the UN IGME-estimated U5MR (IGME 2015). These exclusion rules are based on external information that suggests that some NMR observations are unreliable due to, for example, poor survey quality or under coverage of VR systems. A total of 16% of the 4,678 observations were excluded.

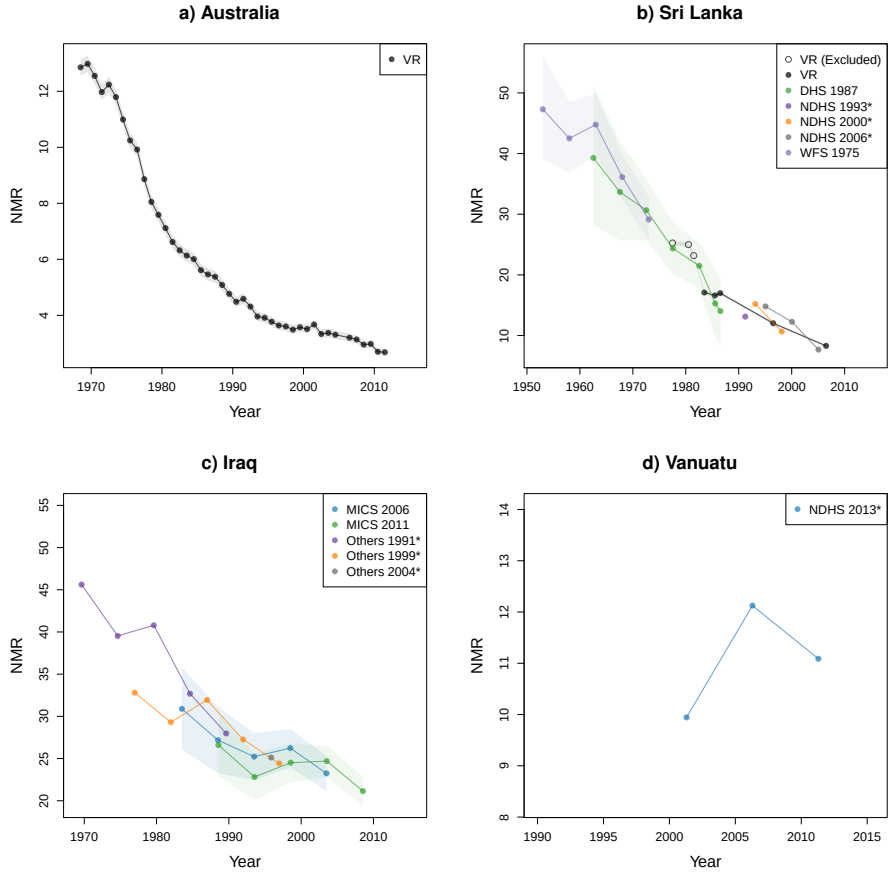
Table 1: Summary of the NMR data availability by source type

Source	Sampling errors	No. of series	No. of countries	No. of obs.	No. of country-years
VR	Calculated	105	105	2607	2607
SVR	Calculated	3	3	79	78
DHS	Reported	239	81	1212	934
DHS	Unreported	16	15	50	48
Other DHS	Reported	52	42	251	251
Other DHS	Unreported	26	21	78	75
MICS	Reported	16	14	81	73
MICS	Unreported	12	12	49	46
Others	Reported	24	16	119	111
Others	Unreported	72	36	152	151

Note: The totals include observations that were excluded from the estimation.

Figure 1 illustrates examples of the data available for four countries. The shaded area around the observations has a width of two times the sampling error (for survey data) or stochastic error (for VR data). The NMR for Australia (Figure 1), as calculated from the full VR data time series, has a trend over time that is relatively regular, and the uncertainty is low. Data for Sri Lanka indicates that the NMR is roughly five times as high as Australia. Data is available from 1950, but the VR data series is incomplete. The rest of the data comes from the WFS, DHS, and National DHS. There are multiple estimates for some years, and the uncertainty around the estimates varies by source and year. The uncertainty around the VR data is much less than for the survey data. The National DHS series does not have estimates of sampling error. Iraq (Figure 1) has no VR data, and the estimates are constructed from MICS and two other surveys: the Infant and Child Mortality and Nutrition Survey, and the Child and Maternal Mortality Survey. Again, there are multiple estimates for some time points, and uncertainty levels and availability vary. Finally, Vanuatu (Figure 1) has only three observation points from one National DHS.

Figure 1: Neonatal mortality data (deaths per 1,000 births) for four selected countries



Note: The different colored circles represent different data sources, as described in the plot legends. An open circle indicates that the observation was excluded from the analysis. The shaded area around the data series represents the stochastic error (in the case of VR data) and sampling error (in the case of survey data). Survey data series that did not have reported sampling errors do not have a shaded region on the plots and are marked with an asterisk (*) in the legends.

3. Method

The aim is to produce estimates of the NMR for all countries in the world and report the associated uncertainty around the estimates. The model needs to be flexible enough to

estimate NMR in a variety of situations, as illustrated in Figure 1. The estimates should follow the data closely for countries with reliable data and low uncertainty. On the other hand, the model estimates need to be adequately smooth in countries with relatively large uncertainty and erratic trajectories. The model also needs to be able to estimate NMR over the period 1990–2015 for all countries, including those countries where there is limited or no data available. To achieve these goals, our proposed model utilizes the relationship between the U5MR and NMR: as the level of U5MR decreases, the proportion of deaths under 5 that are neonatal tends to increase. In addition, the model also allows for country-specific effects and time trends to capture data-driven trends in data-rich countries. The term 'data-driven' refers to a model setup where the NMR estimates over time are explicitly influenced by temporal changes in the NMR data. This model is in contrast to a model where temporal changes in NMR estimates are driven by trends in U5MR only, as was the case with the previous IGME model (Oestergaard et al. 2011).

In the NMR model, we use country-year specific U5MRs as explanatory variables and also to obtain final estimates of NMR. All estimates of U5MR used in the model were obtained from the UN IGME (IGME 2015).

3.1 Model overview

Write $N_{c,t}$ and $U_{c,t}$ as the NMR and U5MR for country c at time t, respectively, with $U_{c,t}$ given by the IGME U5MR estimate for that country-year. Note that $N_{c,t}$ and $U_{c,t}$ are always expressed in units of deaths per 1,000 live births. We explain the model setup in terms of the ratio

$$R_{c,t} = \frac{N_{c,t}}{U_{c,t} - N_{c,t}},$$

which refers to the true ratio of neonatal deaths compared to deaths in months 2 to 60. We constrain $R_{c,t} > 0$ such that $0 \leq \frac{N_{c,t}}{U_{c,t}} \leq 1$ to guarantee that NMR estimates are not greater than U5MR estimates. The true ratio $R_{c,t}$ is modeled as follows:

$$R_{c,t} = f(U_{c,t}) \cdot P_{c,t}, \tag{1}$$

where $f(U_{c,t})$ is the overall expected ratio given the current level of U5MR and $P_{c,t}$ is a country-specific multiplier to capture deviations from the expected relationship.

The observed ratio $r_{c,i}$, which refers to the i-th observation of the ratio in country c, is expressed as a combination of the true ratio and some error, i.e.,

$$\begin{aligned} r_{c,i} &= R_{c,t[c,i]} \cdot \epsilon_{c,i} \\ \implies \log(r_{c,i}) &= \log(R_{c,t[c,i]}) + \delta_{c,i} \\ &= \log f(U_{c,t[c,i]}) + \log P_{c,t[c,i]} + \delta_{c,i} \end{aligned} \tag{2}$$

for $c = 1, 2, \ldots, C$ and $i = 1, \ldots, n_c$, where $C = 195$ (the total number of countries) and n_c is the number of observations for country c. The index $t[c, i]$ refers to the observation year for the i-th observation in country c, $\epsilon_{c,i}$ is the error of observation i, and $\delta_{c,i} = \log(\varepsilon_{c,i})$. Note that throughout the paper, 'log' refers to the natural logarithm.

The following sections explain how we chose to model the expected ratio $f(U_{c,t})$, the country-specific multiplier $P_{c,t}$, and the error term $\delta_{c,i}$. The Appendix details other aspects of the method, including the projection method, estimation for countries with no data, and crisis and HIV/AIDS adjustments.

3.2 Global relationship with U5MR

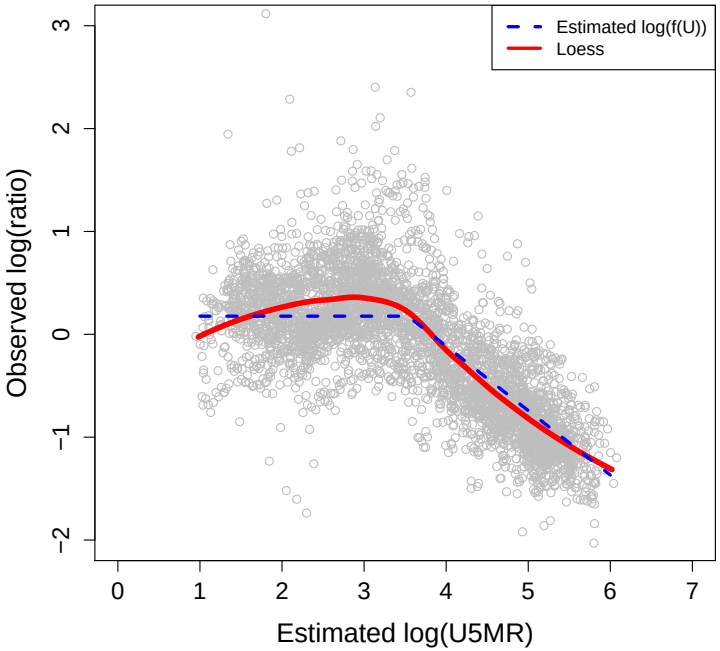
The first step in modeling the ratio of neonatal to non-neonatal deaths is to find an appropriate function $f(U_{c,t})$ in Equation 1, which captures the expected value of the ratio given the current level of U5MR. We modeled $f(U_{c,t})$ on the log scale. Figure 2 shows a scatter plot of log-transformed observed ratios $\log(r_{c,i})$ versus $\log(U_{c,t[c,i]})$. The relationship between the two variables appears to be relatively constant up to around $\log(U_{c,t}) = 3.5$, after which point the log ratio decreases linearly with decreasing $\log(U_{c,t})$. Given this observed relationship, $\log f(U_{c,t})$ is modeled as follows:

$$\log f(U_{c,t}) = \begin{cases} \beta_0 & \text{for } U_{c,t} \leq \theta, \\ \beta_0 + \beta_1 \cdot (\log(U_{c,t}) - \log(\theta)) & \text{for } U_{c,t} > \theta. \end{cases}$$

This implies that, below a cutpoint θ, $\log f(U_{c,t})$ is modeled as a constant β_0. Above the cutpoint, $\log f(U_{c,t})$ is represented linearly as a function of $\log(U_{c,t})$ with slope β_1.

Figure 2 illustrates the fitted relationship between the ratio and the level of U5MR from the NMR model, $\log f(U_{c,t})$. The posterior median estimate for the cutpoint θ is 34.2 deaths per 1,000 births (90% CI: [33.7, 34.5]). At U5MR levels that are higher than θ, the β_1 coefficient suggests that a 1% increase in the U5MR leads to a 0.65% decrease (90% CI: [0.61, 0.71]) in the ratio $R_{c,t}$. The fitted line is quite similar in shape to the loess curve fitted to the data, shown by the red line in Figure 2.

Figure 2: Observed and estimated relation between the ratio of neonatal and non-neonatal deaths and under-5 mortality



Note: Observations $\log r_{c,i}$ are displayed with grey dots and plotted against $\log U_{c,t[c,i]}$. The loess fit to the observations is shown in red, and the estimated relation (function $f(U_{c,t})$) is added in blue (dashed line).

3.3 Country-specific multiplier

This section details how the country-specific multiplier $P_{c,t}$ is modeled. Although there is a relationship between the neonatal ratio and U5MR at the aggregate level, the relationship between $R_{c,t}$ and $U_{c,t}$ is likely to differ by country. For instance, some countries may have higher or lower levels of NMR than what is expected given the level of U5MR. In addition, within a particular country, the relationship between NMR and U5MR may not be constant over time, so the model should be flexible enough to also allow for temporal changes. The purpose of the country-specific term $P_{c,t}$ in Equation 1 is to capture data-driven differences across countries and also within countries over time.

The country-year multiplier $P_{c,t}$ was modeled on the log scale with a basis-splines (B-splines) regression model:

$$\log(P_{c,t}) = \sum_{k=1}^{K_c} B_{c,k}(t)\alpha_{c,k},$$

where $B_{c,k}(t)$ refers to the k-th B-spline function for country c evaluated at time t and $\alpha_{c,k}$ is the k-th splines coefficient for country c. The B-splines $B_{c,k}(t)$, which are illustrated in Figure 3 for Nigeria, were constructed using cubic splines. In the figure, each $B_{c,k}(t)$ is represented in a different color at the bottom. Spline placement is determined by knot points, indicated by gray dotted vertical lines. Knot points occur where the spline function is at its maximum. Country c has a total of K_c knot points defined by $t_1 < t_2 < \cdots < t_{K_c}$. K_c is the number of B-splines needed to cover the period up to 2015 and back to 1990 or the start of the observation period, whichever is earlier. In terms of knot spacing, the same interval length of 2.5 years was used in each country, regardless of the number or spacing of observations. The consistent interval length was chosen to be able to exchange information across countries about the variability in changes between spline coefficients. Knot placement was determined by placing one knot half an interval before the most recent observation year in each country. Because the most recent observation year differs by country, the splines $B_{c,k}(t)$ also differ by country.

Figure 3 also illustrates the fitting of the country-specific multiplier $\log P_{c,t}$ for Nigeria. The splines regression for $\log P_{c,t}$ captures any pattern in the data on the log scale after the global relation between the ratio and U5MR, as expressed by $f(U_{c,t})$, has been taken into account. As such, the y-axis in Figure 3 refers, on the log scale, to the difference between the observed data points and their expected level given the U5MR, i.e., $\log(r_{c,i}) - \log f(U_{c,t[c,i]})$. The different colored dots connected with lines represent residual data points from different sources available for Nigeria. The estimated $\log P_{c,t}$ at a particular time point t is given by a linear combination of the $B_{c,k}$ at point t and the estimated coefficients $\alpha_{c,k}$.

The splines regression model for $\log(P_{c,t})$ is very flexible in order to be able to capture patterns in the data. However, in situations where the data is sparse, limited information on a subset of the spline coefficients $\alpha_{c,k}$ can result in an implausible fit. We impose smoothness on the fits by penalizing differences in adjacent spline coefficients $\alpha_{c,k}$. This is referred to as Penalized splines, or P-splines regression (Eilers and Marx 1996; Currie and Durban 2002).

In the P-splines regression, spline coefficients are modeled as a combination of an overall mean value λ_c and $K_c - 1$ first-order differences

$$\boldsymbol{\varepsilon}_c = (\alpha_{c,2} - \alpha_{c,1}, \alpha_{c,3} - \alpha_{c,2}, \ldots, \alpha_{c,K_c} - \alpha_{c,K_c-1}).$$

The λ_c can be interpreted as a country-specific intercept, representing deviations in the level from the overall global relationship between $R_{c,t}$ and $U_{c,t}$. We model the λ_c's centered at zero:

$$\lambda_c \sim N(0, \sigma_\lambda^2).$$

The ε_c term represents fluctuations around the country-specific intercept. These fluctuation terms allow for the $P_{c,t}$ term to be influenced by the changes in the level of the underlying data. The fluctuations are modeled as

$$\varepsilon_{c,k} \sim N(0, \sigma_{\varepsilon_c}^2). \tag{3}$$

The variance $\sigma_{\varepsilon_c}^2$ essentially acts as a country-specific smoothing parameter. The smoothness of a particular country's trajectory depends on the regularity of the trend in the data and also the measurement errors associated with the data points. As $\sigma_{\varepsilon_c}^2$ decreases, the fluctuations go to zero, and the $\alpha_{c,k}$'s become a country-specific intercept with no change over time. The $\sigma_{\varepsilon_c}^2$'s are modeled hierarchically:

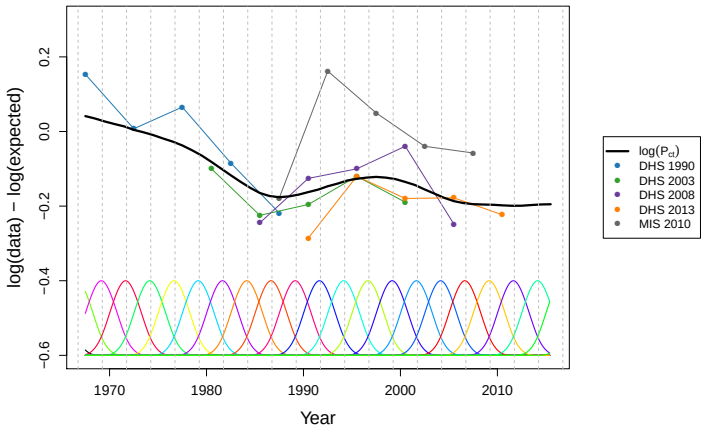
$$\log(\sigma_{\varepsilon_c}^2) \sim N(\chi, \psi_\sigma^2), \tag{4}$$

where e^χ can be interpreted as a 'global smoothing parameter' and ψ_σ^2 reflects the across-country variability in smoothing parameters. The hierarchical structure of the model allows information on the amount of smoothing to be shared across countries. The countries with fewer data points and thus less information about the level of smoothness borrow strength from countries with more observations.

The effect of including the country-specific term is shown in Figure 3 for Nigeria. The available data series are shown by the points, and the shaded area around those points represents their associated sampling error. The blue solid line illustrates the fit from the global relation, i.e., $f(U_{c,t})$. The green line illustrates the global relation and country-specific intercept, i.e., a combination of $f(U_{c,t})$ and λ_c. Note that this line has the same shape as the blue line, but has been lowered. The red solid line shows the final fit after inclusion of the fluctuations, i.e., $f(U_{c,t})$, λ_c and ε_c. This allows the fitted trajectory to be more influenced by the data.

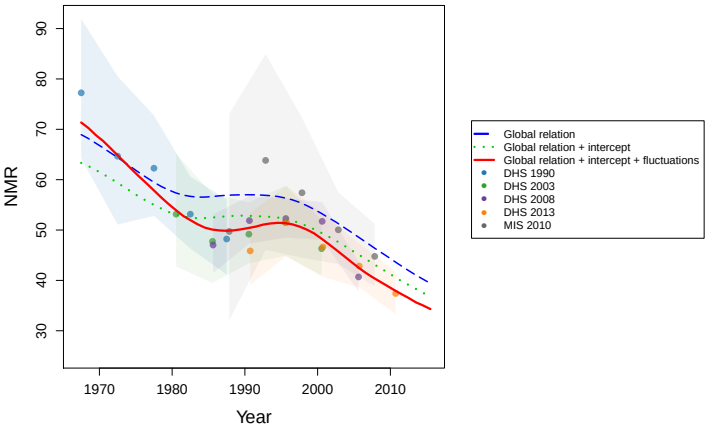
Figure 3: Illustration of splines regression and three fit components for Nigeria

a) Estimate of $\log P_{c,t}$ for Nigeria using splines regression



Note: The y-axis is (on log scale) the difference between the observed data points and the expected level (given by $f(U_{c,t})$). The different colored dots/lines represent data from different sources available. Each basis spline is represented in a different color at the bottom of the figure. These have been scaled vertically for display purposes. The gray dotted vertical lines indicate knot points (every 2.5 years).

b) The three components of fit



Note: The blue dashed line illustrates the fit from the global relation, i.e., $f(U_{c,t})$. The green dashed line illustrates the global relation and country-specific intercept, i.e., a combination of $f(U_{c,t})$ and λ_c. The red solid line shows the final fit after inclusion of the fluctuations i.e., $f(U_{c,t})$, λ_c and $\boldsymbol{\varepsilon}_c$.

The choice to model the country-specific terms $P_{c,t}$ using P-splines regression was motivated by previous work on mortality modeling and forecasting (e.g., Alkema and New 2014, Currie, Durban, and Eilers 2004; D'Amato, Piscopo, and Russolillo 2011). In practice, there are many different ways to model and smooth the country-specific term $P_{c,t}$. For example, $P_{c,t}$ could have been modeled using an autoregressive or autoregressive-moving average (ARMA) process, similar to the modeling approaches for global maternal mortality (Alkema et al. 2016) and contraceptive prevalence (Alkema et al. 2013). However, the use of a splines basis, which results in a regression function that is twice differentiable, gives estimates of $P_{c,t}$ that are relatively smooth compared to an ARMA-based approach. In addition, the P-splines regression approach was chosen for consistency with the current model used by IGME to estimate U5MR.

3.4 Data model

Equation 2 indicates that the observed ratio $r_{c,i}$ is modeled on the log scale as the true ratio $R_{c,t}$ plus some error term $\delta_{c,i}$. We model this error term $\delta_{c,i}$ differently based on the source of the data of the i-th observation. The model imposed on $\delta_{c,i}$ is called the 'data model.'

3.4.1 VR data

For VR data series, the error term $\delta_{c,i}$ is modeled as

$$\delta_{c,i} \sim N(0, \tau_{c,i}^2),$$

where $\tau_{c,i}$ is the stochastic standard error. These are obtained based on standard assumptions about the distribution of deaths in the first month of life. The Appendix gives details. Note that SVR data is modeled the same as VR data, but the $\tau_{c,i}$ term refers to the sampling error.

3.4.2 Non-VR data

For the non-VR data, the error term δ_i is modeled as

$$\delta_{c,i} \sim N(0, \nu_{c,i}^2 + \omega_{s[c,i]}^2),$$

where $\nu_{c,i}$ is the sampling error and $\omega_{s[c,i]}$ is the nonsampling error of the series type s of observation i in country c. Nonsampling error variances are estimated separately for each of the series types listed in Table 1: DHS, other DHS, MICS, and others. The distinction

by series type was made to allow for the possibility that a particular survey may be run in a similar fashion across different countries, and as such may display similar characteristics in terms of nonsampling error.

Sampling error variances were reported for the majority of the non-VR observations (see Table 1). For those observations (c, i) where sampling error was not reported, it was imputed based on the median value of all observed sampling errors of series type $s[c, i]$ within the group-size category of country c. We categorized a country as 'small' if the annual number of births was in the lowest quartile of all countries (corresponding to a maximum of around 25,000 births per year). Table 2 shows the distinction between small and other-sized countries was made due to the differences in observed standard errors. The imputed values for missing standard errors for each size category and series type are shown in Table 2.

Table 2: Values imputed for missing standard errors for survey data by series and country size category

	Country size category	
Series type	**Other**	**Small**
DHS	0.13	0.26
MICS	0.16	0.21
Other DHS	0.14	0.24
Others	0.16	0.22

3.5 Obtaining the final estimates

The model produces estimates of $\log(R_{c,t})$. The corresponding estimate of $N_{c,t}$ is obtained by transforming the ratio and combining it with $U_{c,t}$:

$$N_{c,t} = \text{logit}^{-1}(\log(R_{c,t})) \cdot U_{c,t},$$

because

$$\text{logit}\left(\frac{N_{c,t}}{U_{c,t}}\right) = \log\left(\frac{N_{c,t}}{U_{c,t} - N_{c,t}}\right) = \log(R_{c,t}).$$

The ratio estimates are recombined with IGME estimates of $U_{c,t}$. However, using only the median estimates of $U_{c,t}$ does not take into account the level of uncertainty in the $U_{c,t}$ estimates and correspondingly under-represents the level of uncertainty in the $N_{c,t}$. As such, the $N_{c,t}$ estimates were generated by randomly combining posterior draws of

$\text{logit}^{-1}(\log(R_{c,t}))$ and of $U_{c,t}$. The result is a series of trajectories of $N_{c,t}$ over time. The best estimate is taken to be the median of these trajectories, and the 5th and 95th percentiles are used to construct 90% credible intervals.

3.6 Computation

The Appendix summarizes the hierarchical model. The model was fitted in a Bayesian framework using the statistical software R. Samples were taken from posterior distributions of the parameters via a Markov Chain Monte Carlo (MCMC) algorithm. This was performed by using JAGS software (Plummer 2003).

In terms of computation, three chains with different starting points were run with a total of 20,000 iterations in each chain. Of these, the first 10,000 iterations in each chain were discarded as burn-in, and every tenth iteration after this was retained. Thus 1,000 samples were retained from each chain, meaning that 3,000 samples were retained for each estimated parameter.

Trace plots were checked to ensure adequate mixing and to verify that the chains were past the burn-in phase. Gelman's $\hat{R}$ (Gelman and Rubin 1992) and the effective sample size were checked to ensure a large enough and representative sample from the posterior distribution. The value of $\hat{R}$ for all parameters estimated was less than 1.1.

4. Results

Estimates of NMR were produced for the 195 UN member countries for at least the period 1990–2015, with periods starting earlier if data was available. This section highlights some key results. Results are also compared to those produced by the method previously used by the IGME.

The estimated global relation (Table 3) suggests that the relationship between the ratio and U5MR is constant up to a U5MR of 34.2 (90% CI: [33.7, 34.5]) deaths per 1,000 births, and the ratio of neonatal to other child mortality is constant at around 1.20 (90% CI: [1.03, 1.25]). This is equivalent to saying that the proportion of deaths under-5 that are neonatal is constant at around 54% (90% CI: [50, 55]). Above a U5MR of 34, the estimated coefficient suggests that, at the global level, a 1% increase in the U5MR is associated with a 0.65% (90% CI: [0.61, 0.70]) decrease in the ratio.

Table 3: Estimates for parameters in global relation

	Median	90% CI
β_0	0.18	(0.03, 0.22)
β_1	–0.65	(–0.70, –0.61)
θ	34.2	(33.7, 34.5)

4.1 Results for selected countries

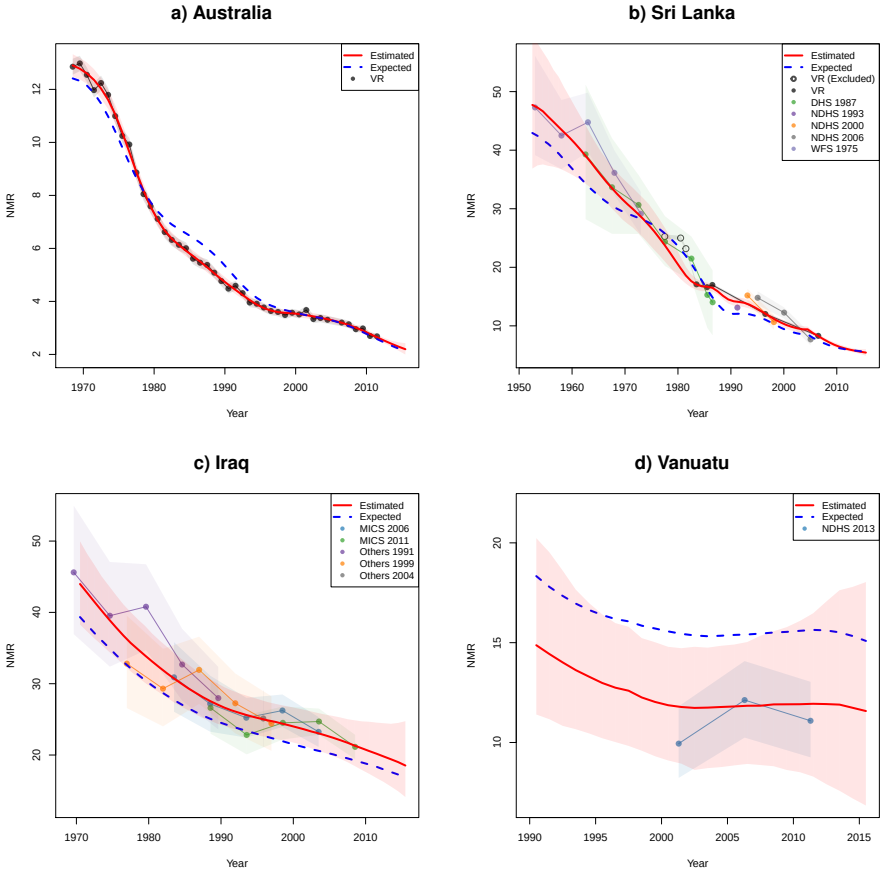
Figure 4 shows the fits for the four countries illustrated in Section 2. In the figures, the blue dashed line represents the expected level of NMR given the country's U5MR. The solid red line and associated shaded area represent model estimates and 90% uncertainty intervals. For Australia (Figure 4), the estimates follow the data closely, given the small uncertainty levels around the data. There has been a steady decrease in NMR since 1970. In earlier time periods, the level of NMR was higher than the expected level (that is, the solid red line is higher than the blue dashed line). This switched in the 1980s and 1990s, and more recently, the estimated and expected levels are close.

For Sri Lanka (Figure 4), the estimates of NMR are informed by the combination of VR and survey data. The VR has a greater influence on the trajectory because of the smaller associated standard errors. In the earlier years, the uncertainty intervals around the estimate are larger due to the higher uncertainty of the data. There is a small spike in the estimate in the year 2004, which is a tsunami-related crisis adjustment.

No VR data was available for Iraq (Figure 4), and the larger sampling errors around the survey data have led to relatively wide uncertainty intervals over the entire period. This is in contrast to Sri Lanka, where uncertainty intervals became narrower when VR data was available. The larger sampling errors in Iraq have also led to a relatively smooth fit (high value of the smoothing parameter), and the shape of the trajectory essentially follows the shape of the expected line.

For Vanuatu (Figure 4), the trajectory is driven by the expected trajectory given Vanuatu's trend in U5MR. The available data determines the country-specific intercept for Vanuatu, which is lower than the expected level. However, the relative absence of data for this country means that the uncertainty around the estimates is high.

Figure 4: Observed and estimated neonatal mortality (deaths per 1,000 births) for selected countries



Note: The blue dashed line represents the expected level of NMR given the country's U5MR. The solid red line and associated shaded area represent model estimates and 90% uncertainty intervals.

4.2 Outlying countries

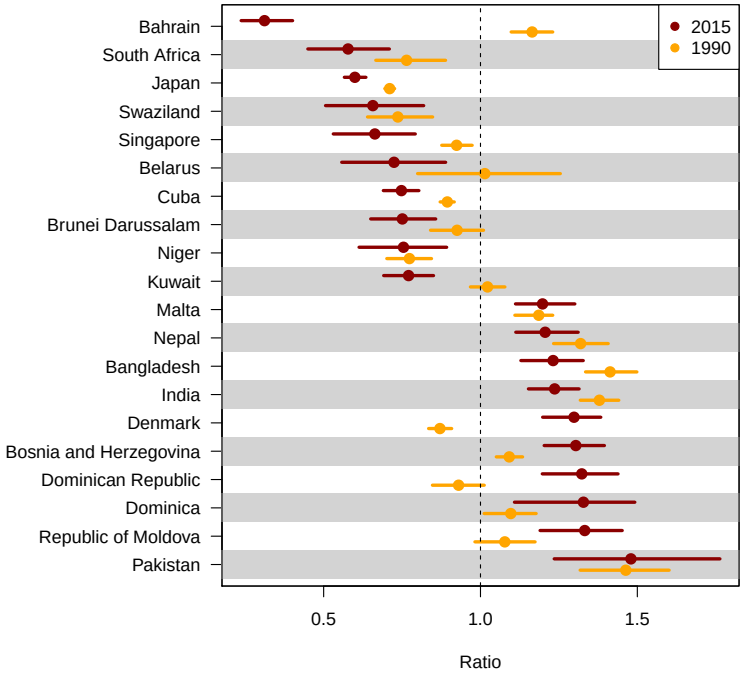
The setup of the model allows for the comparison of the estimated level of NMR to the expected level of NMR given the U5MR. The expected level as predicted by the U5MR is an estimation with $f(U_{c,t})$ only (without the country-specific effect, $P_{c,t}$). We define

a country to be outlying if the estimated NMR in 2015 was higher or lower than the expected level by at least 10%. That is, the ratio of estimated-to-expected was at least 1.1 or less than 0.9 in 2015, and the 95% credible interval does not contain 1. Figure 5 illustrates these countries, and the values of estimated-to-expected in 1990 and 2015.

Countries that have a lower-than-expected NMR include Japan, Singapore, and South Korea, and some African countries such as South Africa and Swaziland. Countries that have a higher-than-expected NMR include several Southern Asian countries, such as Bangladesh, Nepal, India, and Pakistan. The former Yugoslavian countries Croatia, Bosnia and Herzegovina, and Montenegro also have higher-than-expected NMR.

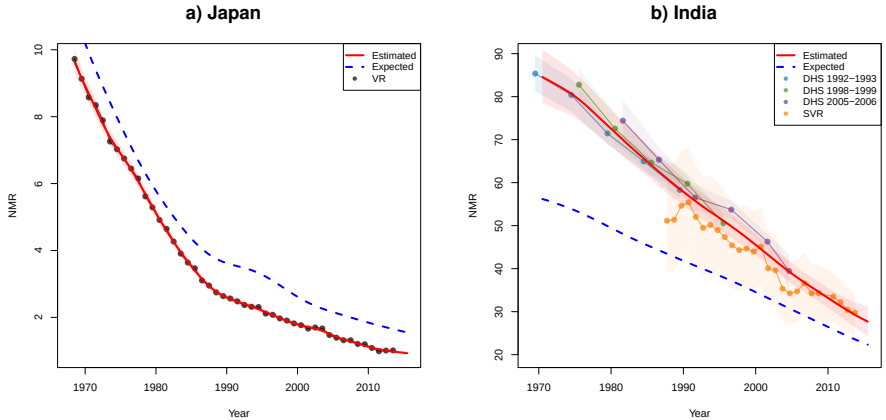
Figure 6 shows estimates through time for two contrasting countries: Japan, which has lower-than-expected NMR, and India, with higher-than-expected NMR. In each of the figures, the red line represents the estimated fitted line (with 90% CIs). The blue line represents the expected level, which can be interpreted as the expected level of NMR in a particular year as predicted by the level of U5MR. The gap between the expected and estimated NMR is being sustained through time for Japan, and has widened since the 1970s. The change in NMR levels for India has been dramatic. Not only is the current NMR around 30% of what it was in 1970, the discrepancy between the expected and estimated levels has decreased through time.

Figure 5: Ratio of estimated NMR to the expected NMR given U5MR for outlying countries



Note: A country is outlying if the estimated NMR in 2015 was higher or lower than the expected level given U5MR by at least 10%. The dot shows the median estimate, and the lines give the 95% uncertainty interval.

Figure 6: Observed and estimated neonatal mortality (deaths per 1,000 births) for selected countries



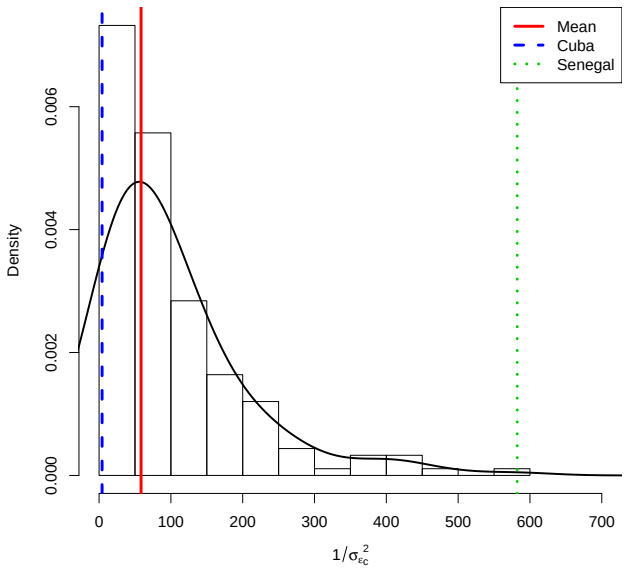
Note: The blue dashed line represents the expected level of NMR given the country's U5MR. The solid red line and associated shaded area represents model estimates and 90% uncertainty intervals.

4.3 Smoothing

The smoothness of the fluctuations, $\sigma^2_{\varepsilon_c}$, is modeled hierarchically, assuming a log-normal distribution with a mean parameter χ (see Equation 4). Smoothing parameters can also be expressed in terms of precision, $1/\sigma^2_{\varepsilon_c}$; Figure 7 shows the distribution of estimated precisions for all countries. The larger the value of the smoothing parameter (precision), the smoother the fit. The estimate of the mean smoothing parameter was around 59 (90% CI: [43, 79]).

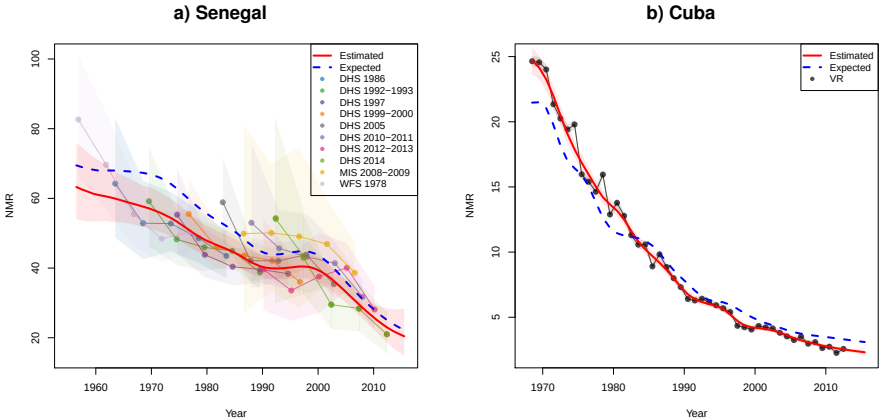
Larger values of smoothing parameters were estimated for countries that had no available VR data but many observations from survey data. Senegal, which had the highest smoothing parameter at a value of 582 (90% CI: [113, 4000]), had a total of 55 observations over a 45-year period (Figure 8). The effect of having many observations with relatively large standard errors is a relatively smooth fit. In contrast, one of the smallest smoothing parameters occurred for Cuba, at around 4 (90% CI: [2, 7]). Cuba is a country with good quality VR data that has relatively small standard errors. This means the fitted line follows the data more closely (Figure 8).

Figure 7: Distribution of estimated precisions ($1/\sigma^2_{\varepsilon_c}$) relating to smoothing parameters for all countries



Note: The red solid line represents the mean value of all precisions. The blue dashed line is the estimated smoothing parameter for Cuba (relatively little smoothing), while the dotted line is the estimated smoothing parameter for Senegal (relatively high smoothing).

Figure 8: Example countries with relatively high smoothing (Senegal) and relatively low smoothing (Cuba)



4.4 Comparison with existing IGME model

It is useful to compare the results of this new model to the NMR results from the model previously used by the IGME. The previous model is described in Oestergaard et al. (2011). In this method, NMR estimates for countries with complete VR series are taken directly from the data. For countries without a complete VR series, a multilevel model is fit using U5MR as a predictor, with a quadratic relationship specified. In addition, the model allows for country-level and region-level random effects:

$$\log(NMR_{c,t}) = \underbrace{\alpha_0 + \beta_1 \log(U_{c,t}) + \beta_2 (\log U_{c,t})^2}_{\log(f(U_{c,t}))} + \underbrace{\alpha_{country[i]} + \alpha_{region[i]}}_{\log(P_{c,t})}$$

For comparison, the new model is:

$$\begin{aligned} \log(R_{c,t}) &= \underbrace{\beta_0 + \beta_1 \cdot (\log(U_{c,t}) - \log(\theta))_{[\log(U_{c,t})>\log(\theta)]}}_{\log(f(U_{c,t}))} \\ &+ \underbrace{\sum_{k}^{K_c} B_{c,k}(t)\alpha_{c,k}}_{\log(P_{c,t})} \end{aligned}$$

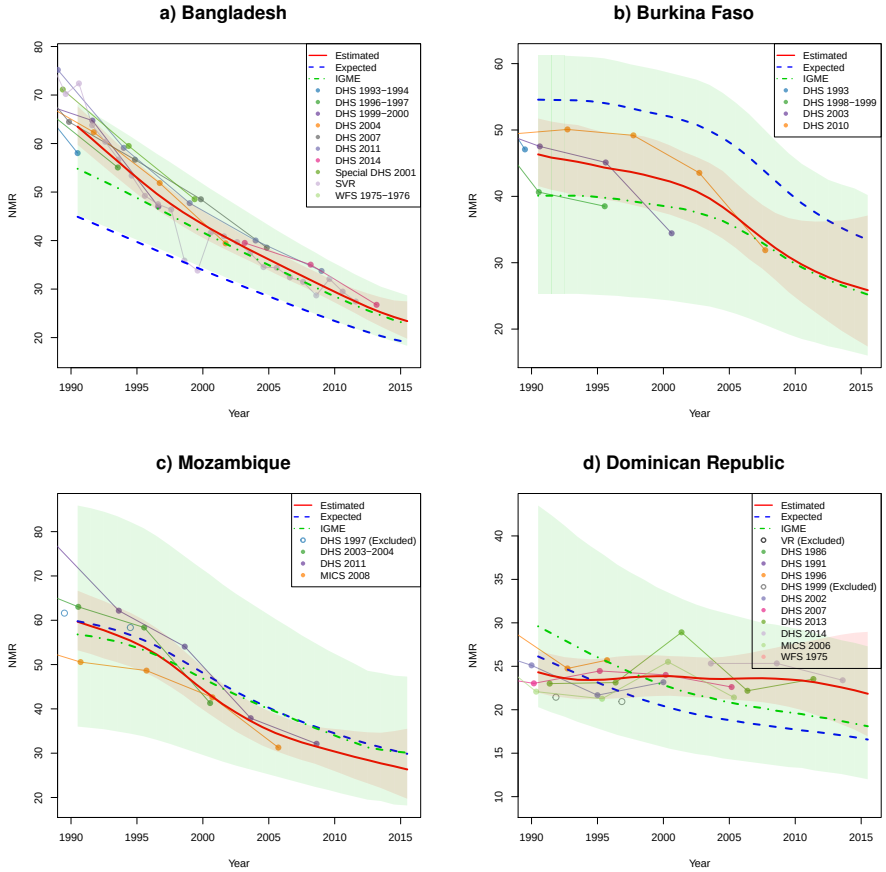
The existing model is similar in that it estimates NMR as a function of U5MR, plus some additional country-specific effect, i.e., a $f(U_{c,t})$ and a $P_{c,t}$. However, one of the main differences between the two models is that for countries with non-VR data, estimates from the new model can be driven by the data, while the previous model is restricted to follow the trajectory of the U5MR in a particular country, plus or minus some country-specific intercept. Table 4 highlights other differences between the two models.

Table 4: Comparison of two models

IGME 2014	New model
Model used for non-VR countries	Model used for all countries
Model relation between NMR and U5MR	Model relation between ratio and U5MR
$f(U_{c,t})$ is quadratic	$f(U_{c,t})$ is linear with changing slope
$P_{c,t}$ is a country and region-specific intercept	$P_{c,t}$ is a country-specific intercept + fluctuations
Country-specific effect constant over time	Country-specific effect can change over time
Only considers sampling error	Data model with sampling and nonsampling error

Figure 9 compares the results of four countries to the estimates from the current IGME model. The estimates from the previous IGME model generally follow the same trajectory as the expected line, as determined by U5MR patterns, and is shifted up or down depending on the estimate of the country-specific effect. In contrast, the estimates from the new model follow the data more closely. The fluctuation part of the country-specific multiplier, $P_{c,t}$, allows the estimated line to move above or below the expected line, as is the case with the Dominican Republic (Figure 9). In addition, there is generally less uncertainty around the estimates in the new model, especially in periods where there is data.

Figure 9: Estimated NMR (deaths per 1,000 births) for four example countries; new model vs. IGME 2014 model



Note: The green dashed line and associated shaded area show the model estimates from the 2014 IGME model.

4.5 Model validation

We assessed model performance through an out-of-sample model validation exercise. In creating a training dataset, rather than removing observations at random, we chose the process of removing data to emulate the way in which new data may be received (Alkema, Wong, and Seah 2012). Mortality databases are updated at least once a year as more data

becomes available. These updates may include not only data for the most recent time period but may also include, for example, retrospective estimates from a survey. Ideally the model should not be sensitive to updates of historical data, so estimates do not change from year to year.

We constructed the training set by leaving out the most recent survey data series, and for countries with only one series (including VR countries), the most recent 20% of data observations were removed. The resulting training dataset was made of around 80% of the total data available.

For the left-out observations, the absolute relative error is defined by

$$e_i = \frac{|r_{c,i} - \tilde{r}_{c,i}|}{\tilde{r}_{c,i}},$$

where $\tilde{r}_{c,i}$ denotes the posterior median of the predictive distribution for a left-out observation $r_{c,i}$ based on the training set. Coverage is defined by

$$\frac{1}{N} \sum 1[r_{c,i} \geq l_{c[i]}(t[i])]1[r_{c,i} < r_{c[i]}(t[i])],$$

where N is the total number of left-out observations considered and $l_{c[i]}(t[i])$ and $r_{c[i]}(t[i])$ the lower and upper bounds of the predictions intervals for the i-th observation. Coverage at the 80%, 90%, and 95% levels was considered.

The validation measures were calculated for 100 sets of left-out observations, where each set consisted of a random sample of one left-out observation per country. Table 5 shows the median and standard deviation of each validation measure. The median absolute relative error between the observations and estimated value was less than 10%, and the coverage of the prediction intervals is approximately as expected.

Table 5: Validation measures, left-out data

	Expected	Median	Std. Dev
Mean absolute relative error	–	0.09	0.02
80% coverage	0.80	0.84	0.02
90% coverage	0.90	0.92	0.02
95% coverage	0.95	0.96	0.02

A similar set of validation measures was calculated comparing the model estimates based on the training dataset with the model estimates based on the full dataset. The absolute relative error is defined as

$$e_i = \frac{|r^*_{c,i} - \tilde{r}_{c,i}|}{r^*_{c,i}},$$

where $r^*_{c,i}$ and $\tilde{r}_{c,i}$ denote the posterior median of the predictive distribution for the observation $r_{c,i}$ based on estimates using the full and training dataset, respectively. Coverage refers to what proportion of the posterior median estimates from the training dataset fall within the 80%, 90%, and 95% bounds of estimates using the full dataset.

Results in Table 6 are reported for estimates up to (and including) 2005, and post 2005. Model performance is better prior to 2005. This is due to the most recent data being removed, so the data prior to 2005 would be very similar between training and test sets. However, the post-2005 measures show that estimates are reasonably consistent between the reduced and full datasets.

Table 6: Validation measures, model comparison

	Expected	**$\leq$ 2005**	**$>$ 2005**
Mean absolute relative error	–	0.05	0.09
80% coverage	$\geq$ 0.80	0.90	0.77
90% coverage	$\geq$ 0.90	0.94	0.84
95% coverage	$\geq$ 0.95	0.96	0.90

5. Discussion

A new model was introduced for estimating NMR. The model can be expressed as the product of an overall relationship with U5MR and a country-specific effect. The overall relationship with U5MR is a simple linear function, while the country-specific effect is modeled through P-spline regression as a country-specific intercept plus fluctuations around that intercept.

Estimates of the NMR were produced for 195 countries, spanning at least the period 1990–2015. The model appears to perform well in a wide variety of situations where the extent and type of data available varies. In many developed countries, where VR data series are complete and uncertainty around the data is low, NMR estimates follow the data closely. On the other hand, where there is limited data available or if uncertainty around the data is high, estimates are more influenced by the trends in U5MR.

The model was fit within a Bayesian hierarchical framework, allowing information about trends in NMR to be exchanged across countries. Through the hierarchical structure, the smoothness in trends in NMR for countries with little data available, or highly uncertain data, is partially informed by countries with more reliable data. While this setup has the potential to introduce biases in country-specific estimates, it allows estimates with reasonable amounts of uncertainty to be produced for all countries, even in the absence of reliable data. Validation exercises did not highlight any problems with bias, suggest-

ing that estimates and uncertainty produced by the model are reliable in a variety of data situations.

We compared model estimates to estimates from the existing IGME model. The notable advantage of this model is that trends in NMR for countries without VR data are driven by the data itself, rather than just reflecting trends in U5MR, as is the case with the existing model. Another advantage of this model is that it is along the same methodological lines as the current model used by IGME to estimate U5MR (Alkema and New 2014). Estimates produced by this model will help to monitor a country's progress in reducing neonatal mortality and reaching the targets set in the SDG.

There are several avenues worth investigating in further research. The choice of a linear function with changing slope for $f(U_{c,t})$ was a data-driven decision, based on the observed relationship in Figure 2. It would be interesting to compare the performance of models that have a functional form that draws upon existing demographic models. For example, an extended version of the Brass relational logistic model (Brass 1971) and Siler models (Siler 1983) can be used to predict survival in the first months of life as a function of the survivorship at older ages.

The potential for bias in estimates from survey data is always a concern. Bias may occur from interviewing a sample that is not representative of the overall population, from selective omission of answers, and can even be influenced by the length of the survey administered (Bradley 2015). The data model included an estimation of an overall level of nonsampling error for each survey type, which may account for some random reporting errors. However, there is scope to further extend the data model to try to better estimate potential bias in survey data estimates.

The focus of this paper was on the methodology. Future work will also focus on interpretation of results. More investigation is needed on what is potentially causing NMR to be higher or lower than expected in outlying countries and whether these are real effects or artifacts of data issues. This distinction is an important one and will become even more so as the focus on child mortality continues to shift toward the early months of life.

References

Alkema, L., Chou, D., Hogan, D., Zhang, S., Moller, A.B., Gemmill, A., Fat, D.M., Boerma, T., Temmerman, M., Mathers, C., and Say, L. (2016). Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group. *The Lancet* 387(10017): 462–474. doi:10.1016/S0140-6736(15)00838-7.

Alkema, L., Kantorova, V., Menozzi, C., and Biddlecom, A. (2013). National, regional, and global rates and trends in contraceptive prevalence and unmet need for family planning between 1990 and 2015: A systematic and comprehensive analysis. *The Lancet* 381(9878): 1642–1652. doi:10.1016/S0140-6736(12)62204-1.

Alkema, L. and New, J.R. (2014). Global estimation of child mortality using a Bayesian B-spline bias-reduction model. *The Annals of Applied Statistics* 8(4): 2122–2149.

Alkema, L., Wong, M.B., and Seah, P.R. (2012). Monitoring progress towards Millennium Development Goal 4: A call for improved validation of under-five mortality rate estimates. *Statistics, Politics and Policy* 3(2): 1–19.

Bhutta, Z.A., Chopra, M., Axelson, H., Berman, P., Boerma, T., Bryce, J., Bustreo, F., Cavagnero, E., Cometto, G., Daelmans, B., de Francisco, A., Fogstad, H., Gupta, N., Laski, L., Lawn, J., Maliqi, B., Mason, E., Pitt, C., Requejo, J., Starrs, A., Victora, C.G., and Wardlaw, T. (2010). Countdown to 2015 decade report (2000–10): Taking stock of maternal, newborn, and child survival. *The Lancet* 375(9730): 2032–2044. doi:10.1016/S0140-6736(10)60678-2.

Bijak, J. and Bryant, J. (2016). Bayesian demography 250 years after Bayes. *Population Studies* 70(1): 1–19. doi:10.1080/00324728.2015.1122826.

Bradley, S. (2015). *More questions, more bias? An assessment of the quality of data used for direct estimation of infant and child mortality in the Demographic and Health Surveys*. Paper presented at the Annual Meeting of the Population Association of America, San Diego, April 30–May 2, 2015.

Brass, W. (1971). On the scale of mortality. In: Brass, W. (ed.). *Biological aspects of demography*. London: Taylor and Francis: 69–110.

Centre for Research on the Epidemiology of Disasters (CRED) (2012). EM-DAT: The CRED international disaster database. [electronic resource]. Brussels: Centre for Research on the Epidemiology of Disasters. www.emdat.be/database.

Currie, I.D. and Durban, M. (2002). Flexible smoothing with P-splines: A unified approach. *Statistical Modelling* 2(4): 333–349. doi:10.1191/1471082x02st039ob.

Currie, I.D., Durban, M., and Eilers, P.H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4(4): 279–298. doi:10.1191/1471082X04st080oa.

D'Amato, V., Piscopo, G., and Russolillo, M. (2011). The mortality of the Italian population: Smoothing techniques on the Lee–Carter model. *Annals of Applied Statistics* 5(2A): 705–724. doi:10.1214/10-AOAS394.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science* 11(2): 89–121. doi:10.1214/ss/1038425655.

Foreman, K.J., Lozano, R., Lopez, A.D., and Murray, C.J. (2012). Modeling causes of death: An integrated approach using codem. *Population Health Metrics* 10(1): 1. doi:10.1186/1478-7954-10-1.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4): 457–472. doi:10.1214/ss/1177011136.

Girosi, F. and King, G. (2008). *Demographic forecasting*. Princeton: Princeton University Press.

Lawn, J.E., Cousens, S., and Zupan, J. (2004). Four million neonatal deaths: When? Where? Why? *The Lancet* 365(9462): 891–900. doi:10.1016/S0140-6736(05)71048-5.

Lozano, R., Wang, H., Foreman, K.J., Rajaratnam, J.K., Naghavi, M., Marcus, J.R., Dwyer-Lindgren, L., Lofgren, K.T., Phillips, D., and Atkinson, C. (2011). Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: An updated systematic analysis. *The Lancet* 378(9797): 1139–1165.

Mahy, M. (2003). Measuring child mortality in AIDS-affected countries. [electronic resource]. New York: United Nations. www.un.org/esa/population/publications/adultmort/UNICEF_Paper15.pdf.

Oestergaard, M.Z., Inoue, M., Yoshida, S., Mahanani, W.R., Gore, F.M., Cousens, S., Lawn, J.E., Mathers, C.D., on behalf of the United Nations Inter-Agency Group for Child Mortality Estimation, and the Child Health Epidemiology Reference Group (2011). Neonatal mortality levels for 193 countries in 2009 with trends since 1990: A systematic analysis of progress, projections, and priorities. *PLOS Medicine* 8(8): 1–13. doi:10.1371/journal.pmed.1001080.

Pedersen, J. and Liu, J. (2012). Child mortality estimation: Appropriate time periods for child mortality estimates from full birth histories. *PLoS Medicine* 9(8): e1001289. doi:10.1371/journal.pmed.1001289.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Hornik, K., Leisch, F., and Zeileis, A. (eds.). *Proceedings of the*

Third International Workshop on Distributed Statistical Computing. Vienna: Technische Universität Wien: 1–9.

Price, M., Klingner, J., and Ball, P. (2013). Preliminary statistical analysis of documentation of killings in Syria. [electronic resource]. Palo Alto: Benetech Human Rights Program. www.ohchr.org/Documents/Countries/SY/PreliminaryStatAnalysisKillingsInSyria.pdf.

Raftery, A.E., Li, N., Ševčíková, H., Gerland, P., and Heilig, G.K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences* 109(35): 13915–13921. doi:10.1073/pnas.1211452109.

Schmertmann, C., Zagheni, E., Goldstein, J.R., and Myrskylä, M. (2014). Bayesian forecasting of cohort fertility. *Journal of the American Statistical Association* 109(506): 500–513.

Siler, W. (1983). Parameters of mortality in human populations with widely varying life spans. *Statistics in Medicine* 2(3): 373–380.

United Nations AIDS (UNAIDS) (2015). UNAIDS gap report. [electronic resource]. Geneva: United Nations AIDS. www.unaids.org/en/resources/campaigns/2014/2014gapreport/gapreport.

United Nations Inter-Agency Group for Child Mortality Estimation (IGME) (2015). Levels and trends in child mortality: Report 2015. [electronic resource]. New York: United Nations Inter-Agency Group for Child Mortality Estimation.

United Nations (UN) (2015). Millennium development goal progress report. [electronic resource]. Geneva: United Nations. www.un.org/millenniumgoals/news.shtml.

United Nations (UN) (2017). Health: United Nations sustainable development. [electronic resource]. Geneva: United Nations. www.un.org/sustainabledevelopment/health.

Walker, N., Hill, K., and Zhao, F. (2012). Child mortality estimation: Methods used to adjust for bias due to AIDS in estimating trends in under-five mortality. *PLOS Medicine* 9(8): 1–7. doi:10.1371/journal.pmed.1001298.

World Health Organization (WHO) (2013). WHO methods and data sources for global causes of death 2000–2011. Geneva: World Health Organization. (Global Health Estimates Technical Paper WHO/HIS/HSI/GHE/2013.3). www.who.int/healthinfo/statistics/GHE_TR2013-3_COD_MethodsFinal.pdf.

You, D., Hug, L., Ejdemyr, S., Idele, P., Hogan, D., Mathers, C., Gerland, P., New, J.R., and Alkema, L. (2015). Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic

analysis by the UN Inter-Agency Group for Child Mortality Estimation. *The Lancet* 386(10010): 2275–2286. doi:10.1016/S0140-6736(15)00120-8.

Appendix

Model summary

The full model is summarized below.

$$
\begin{aligned}
r_{c,i} &\sim N(R_{c,t[c,i]}, \delta_i^2) \\
\delta_i^2 &= \begin{cases} \tau_{c,i}^2 & \text{for VR and SVR data,} \\ \nu_{c,i}^2 + \omega_{s[c,i]}^2 & \text{for non-VR data} \end{cases} \\
R_{c,t} &= f(U_{c,t}) \cdot P_{c,t} \\
\log(f(U_{c,t})) &= \beta_0 + \beta_1 \cdot (\log(U_{c,t}) - \log(\theta))_{[U_{c,t}>\theta]} \\
\log(P_{c,t}) &= \sum_{k=1}^{K_c} B_k(t)\alpha_{c,k} \\
\alpha_{c,k} &= \lambda_c + [\boldsymbol{D}'_{K_c}(\boldsymbol{D}_{K_c}\boldsymbol{D}'_{K_c})^{-1}\boldsymbol{\varepsilon}_c]_k \\
\lambda_c &\sim N(0, \sigma_\lambda^2) \\
\varepsilon_{c,q} &\sim N(0, \sigma_{\varepsilon_c}^2) \\
\log(\sigma_{\varepsilon_c}^2) &\sim N(\chi, \psi^2),
\end{aligned}
$$

where

- $R_{c,t}$ is the true ratio in country c at time t, $R_{c,t} = \frac{N_{c,t}}{U_{c,t}-N_{c,t}}$, where N_{ct} and $U_{c,t}$ are the NMR and U5MR for country c at time t, respectively.
- $r_{c,i}$ is observation i of the ratio in country c.
- $\tau_{c,i}$ is the stochastic standard error, $\nu_{c,i}$ is the sampling error, and $\omega_{s[c,i]}^2$ is non-sampling error for series type s.
- β_0 is the global intercept, β_1 is the global slope with respect to U5MR, θ is the level of U5MR at which β_1 begins to act.
- $P_{c,t}$ is a country-specific multiplier for country c at time t.
- $B_k(t)$ is the kth basis spline evaluated at time t and $\alpha_{c,k}$ is splines coefficient k.
- λ_c is the splines intercept for country c.
- $\boldsymbol{D}_{K_c}$ is a $K_c \times (K_c - 1)$ first-order difference matrix: $D_{K_c i,i} = -1$, $D_{K_c i,i+1} = 1$ and $D_{K_c i,j} = 0$ otherwise.
- $\varepsilon_{c,q}$ are fluctuations around the country-specific intercept.
- $\sigma_{\varepsilon_c}^2$ is the country-specific smoothing parameter, modeled hierarchically on the log-scale with mean χ and variance ψ^2.

The model was fit in a Bayesian framework. Priors are given by

$$\begin{aligned}
\omega &\sim U(0,40) \\
\beta_0 &\sim N(0,100) \\
\beta_1 &\sim N(0,100) \\
\theta &\sim U(0,500) \\
\sigma_{\lambda c} &\sim U(0,40) \\
\chi &\sim N(0,100) \\
\psi &\sim U(0,40).
\end{aligned}$$

Other aspects of the method

Stochastic errors for the VR model

Recall that the observed ratio $r_{c,i}$, which refers to the i-th observation of the ratio in country c, is expressed as a combination of the true ratio and some error, i.e.,

$$\begin{aligned}
r_{c,i} &= R_{c,t[c,i]} \cdot \epsilon_{c,i} \\
\implies \log(r_{c,i}) &= \log(R_{c,t[c,i]}) + \delta_{c,i}
\end{aligned} \tag{5}$$

for $c = 1, 2, \ldots, C$ and $i = 1, \ldots, n_c$, where $C = 195$ (the total number of countries) and n_c is the number of observations for country c. The index $t[c,i]$ refers to the observation year for the i-th observation in country c, $\epsilon_{c,i}$ is the error of observation i, and $\delta_{c,i} = \log(\varepsilon_{c,i})$.

For VR data series, the error term $\delta_{c,i}$ is modeled as

$$\delta_{c,i} \sim N(0, \tau^2_{c,i}),$$

where $\tau^2_{c,i}$ is the stochastic standard error. These can be obtained once some standard assumptions are made about the distribution of deaths in the first month of life. We assume that deaths before age 5 d_5 are distributed

$$d_5 \sim Pois(B \times {}_5q_0),$$

where B is live births and ${}_5q_0$ is the probability of death between ages 0 and 5. Additionally, we assume that deaths in the first month of life d_n are distributed

$$d_n \sim Bin(d_5, p),$$

where $p = {}_nq_0/{}_5q_0$ and ${}_nq_0$ is the probability of death in the first month of life. Note that the values of ${}_nq_0$ and ${}_5q_0$ come from the raw data.

The stochastic error was obtained via simulation. For each year corresponding to observation i in country c,

- a total of 3,000 simulations of under-5 deaths d_5 were drawn from a Poisson distribution $d_5^{(s)} \sim Pois(B \times {}_5q_0)$;
- a total of 3,000 simulations of neonatal deaths d_n were drawn from a Binomial distribution $d_n^{(s)} \sim Bin(d_5^{(s)}, p)$;
- the ratio $y^{(s)} = logit\left(\frac{d_n^{(s)}}{d_5^{(s)}}\right)$ was calculated for each of the simulated samples and the standard error $\tau_{c,i}$ was calculated as $\sigma(\boldsymbol{Y})$ where $\boldsymbol{Y} = (y^{(1)}, y^{(2)}, ...y^{(s)})$, s = 3,000.

It is possible that the stochastic variation as assessed in this simulation approach underestimates the true stochastic uncertainty. We used the results from the validation exercise described in Section 4.5 to determine the coverage at 80%, 90%, and 95% uncertainty levels for VR data only. At all levels the actual coverage level was at least as big as the nominal coverage level. This suggests that either (1) the simulation setup was sufficient to capture stochastic variation in the neonatal deaths and deaths below age 5, or (2) that any underestimation of stochastic uncertainty is compensated by an overestimation of uncertainty associated with the true ratio. Hence, the validation suggests that credible intervals for the ratio are either well calibrated or conservative, which is preferable to underestimating the uncertainty associated with the true outcome.

SVR data

For SVR data, the value for the sampling error was imputed based on the sampling error for $U_{c,t[c,i]}$ SVR data, and the observed ratio between the stochastic error of $r_{c,i}$, and the stochastic error of $U_{c,t[c,i]}$. On average, the stochastic error of $r_{c,i}$ was twice as large as the stochastic error of $U_{c,t[c,i]}$. In addition, the sampling error for $U_{c,t[c,i]}$ SVR data was assumed to be 10%. As such, a value of 20% was imputed for the sampling error for $r_{c,i}$ SVR data.

Projection

When producing NMR estimates, generally data is not available up to the most recent year of interest for the majority of countries, and countries may have longer series of

missing data. As such, country trajectories needed to be projected forward to the year 2015.

The parameters β_0, β_1, and θ, which make up the expected relation with $U_{c,t}$, are fixed over time, as is the country-specific intercept, λ_c. The component that needs to be projected is the random fluctuations part. These $\varepsilon_{c,k}$ were assumed to be normally distributed around zero, with some variance $\sigma^2_{\varepsilon_c}$ (Equation 3). This assumption is used to project the $\varepsilon_{c,k}$ (and thus the splines).

Start at the first $\alpha_{c,k}$ that is past the last year of observed data. For each time period to be projected:

- Draw $\varepsilon_{c,k} \sim N(0, \sigma^2_{\varepsilon_c})$ to obtain $\alpha_{c,k} = \varepsilon_{c,k} + \alpha_{c,k-1}$
- Repeat to generate α_k for k up to K_c, where K_c is the number of knots needed to cover the period up to 2015.

The simulated $\varepsilon_{c,k}$ are generally close to zero, so the method essentially propagates the level of the most recent $\alpha_{c,k}$ that overlaps with the data period with the slope of the expected trajectory, as determined by $f(U_{c,t})$. The projection exercise is necessary in order to maintain a consistent level of uncertainty in the estimates.

Recalculation of VR data for small countries

Several island nations and other small countries have vital registration data available to calculate NMR. However, observations from these small countries are prone to large stochastic error, which can create erratic trends in NMR over time.

To help avoid this issue, observations from adjacent time periods in a particular country are recombined if the coefficient of variation of the observation is greater than 10%. The result is a smaller set of observations with smaller standard errors that display a smoother trend. Figure A-1 shows the example of Saint Vincent and the Grenadines on which this process was applied.

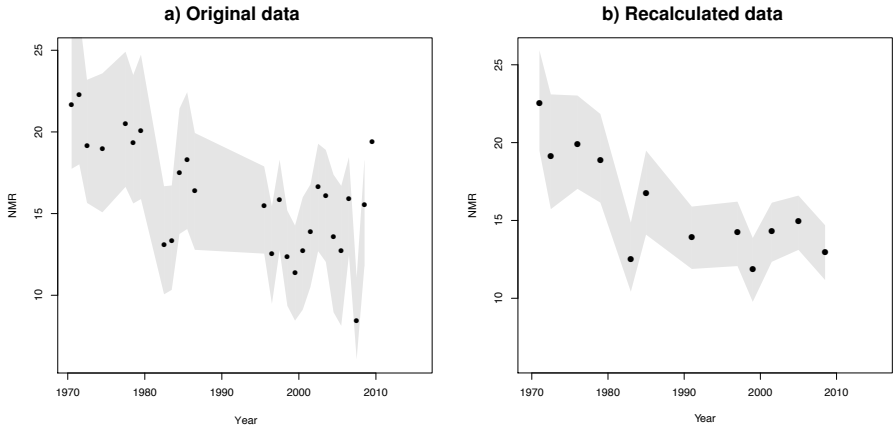
NMR is recalculated using the original NMR observations and annual number of live births. For two adjacent years that are to be recalculated,

- the number neonatal deaths in each year is first calculated as NMR $\times$ live births;
- the combined NMR for the two years is then the total neonatal deaths divided by total births over the two years;
- the standard error of the new NMR estimate is then recalculated based on the process described in 3.4.1.

After recalculation, the coefficient of variation is calculated for the new estimate.

If it is still greater than 10%, the NMR is recalculated again, recombining with the previous adjacent year.

Figure A-1: Recalculation of VR data: Saint Vincent and the Grenadines



Crisis deaths

For some countries, there are known natural or political crises that have caused an excess of deaths; for example, the Rwandan genocide or, more recently, the Haiti earthquake and conflict in Syria. For the crisis years, the survey data is unlikely to be representative of the actual number of deaths.

Adjustments were made to the relevant crisis country-years, using estimates compiled by the World Health Organization (WHO). The WHO uses external data sources on the number of deaths, including the Centre for Research on the Epidemiology of Disasters International Disaster Database (CRED 2012) and estimates from the UN Office of the High Commissioner for Human Rights for the Syrian conflict (Price, Klingner, and Ball 2013). The WHO estimates the proportion of deaths that occur under the age of 5 (WHO 2013). From there, the best guess of the number of crisis deaths that occur within the first month is simply 1/60th of the total deaths under 5 years.

Estimation of crisis countries was firstly done without any crisis adjustments. In addition, the global relation with U5MR, $f(U_{c,t})$, is fit to crisis-free $U_{c,t}$ estimates. The relevant adjustments to country-years were then made post estimation. This was to ensure that the crisis deaths, which are specific to particular years, do not have an effect on the splines estimation.

HIV/AIDS countries

Although there have been vast improvements in recent years, many countries in sub-Saharan Africa still suffer from relatively high levels of HIV/AIDS-related deaths. This has a substantial effect on the child mortality – if children living with HIV are not on antiretroviral treatment, a third will not reach their first birthday, and half will not reach their second birthday (UNAIDS 2014). However, it is unlikely that children with HIV will die within the neonatal period, and so HIV/AIDS itself does not have an explicit effect on the NMR (although there may be indirect effects on mortality, for example through losing their mother to HIV) (Mahy 2003).

Due to this disproportionate effect of HIV/AIDS on U5MR compared to NMR, there are several adjustments made to the inputs used in the model, which leads to NMR being modeled as a function of 'HIV-free' U5MR. Firstly, the U5MR data used in the ratio observations is adjusted to incorporate reporting bias. This adjustment accounts for the higher maternal mortality among HIV-positive mothers, which leads to under-estimation of U5MR from surveys (Walker, Hill, and Zhao 2012). Once adjusted, the AIDS deaths are removed from U5MR, using estimates of deaths provided by UNAIDS (UNAIDS 2014). The result is a ratio of neonatal to other child mortality which is free of AIDS deaths. In addition, the global relation with U5MR, $f(U_{c,t})$, is fit to AIDS-free U5MR. Unlike the crisis adjustments, no AIDS deaths were added in post-estimation, because it is assumed no neonatal deaths are due to HIV/AIDS.

Countries with no data

There were twelve UN-member countries for which the IGME produces NMR estimates, but where there is no available data. For these countries, the estimates of NMR are based on the global relation with U5MR, $f(U_{c,t})$. Additionally, some steps are needed to obtain the appropriate uncertainty around these estimates. For country c

- draw $\lambda_c \sim N(0, \sigma_\lambda^2)$;
- set $\alpha_1 = \lambda_c$;
- draw $\varepsilon_1 \sim N(0, \sigma_\varepsilon)$; where $\sigma_\varepsilon = e^\chi$ is the global smoothing parameter, based on equation 4;
- set $\alpha_2 = \alpha_1 + \varepsilon_1$;
- repeat to generate α_k for $k = 3, \ldots, K_c$. K_c is the number of spline knots needed to cover the period 1990–2015.